Final Project

Panko Aliaksandr

December 13, 2017

Contents

1 Project Objectives

The idea of the project is to check common assumptions regarding stock markets:

- Stock returns follow a normal distribution
- Stock prices follow a lognormal distribution
- Efficient market hypothesis

Knowledge of distribution is required to make correct predictions.

2 Methodology

Having historical data one can compare real data with the data from a distribution with estimated parameters. Using hypothesis testing one can conclude how likely that the historical data fits the distribution.

2.1 Stock returns distribution

It is common assumption that stock returns follow a normal distribution. In this case normality test can be executed to check a hypothesis that data follows normal distribution. The conclusion is made based on probability of the hypothesis acceptance, which is reflected by p-value.

2.2 Stock prices distribution

It is common assumption that stock prices follow a lognormal distribution. In this case lognormality test can be executed to check a hypothesis that data follows lognormal distribution. The conclusion is made based on probability of the hypothesis acceptance, which is reflected by p-value.

2.3 Market efficiency

In this part of the project theoretical probability of Black Monday is computed. The probability is expected to be very low, however history shows that such events are real. Theoretically, prices that follows random walk must be distributed normally with annual volatility 20% and mean 0 (this is given fact). These parameters are used during probability estimation.

3 Project Algorithm

1. Data Download
2. Returns calculation for the indices
3. Normality tests for all indices
4. Lognormality tests for all indices
5. Statistics analysis
6. Black Monday probability calculation
7. Compare lognormal model with fractals

4 Class Structure

My project has 3 classes:

1. Index class
2. Indices analysis class
3. Probability test class

4.1 Index class

Index class has next attributes:

- Index name
- Analyzing period
- Historical values
- Returns

The class has next methods:

- Download historical data
- Calculate daily returns
- Perform normality test
- Perform lognormality test
- Set of getters

4.2 Indices analysis class

The class provides general methods to work with a list of indices. The class has next attributes:

- List of indices
- Indices names
- Analyzing period

The class has next methods:

- Download historical data
- Perform normality test
- Plot indices values distribution
- Plot indices returns distribution
- Set of getters

4.3 Probability test

The class provides gives the probability of an event.

- Event
- Expected value of an distribution
- Volatility

The class has next methods:

- Check probability
- Set of getters

5 Implementation Details

This section devoted to the most important python tools used during the project

5.1 Data Download Process

The data was downloaded from Yahoo! Finance using $pdr.get_data\_yahoo$ ()

5.2 Returns Calculation

Daily returns are computed using the formula:

$$daily\_return(in\%) = \frac{today\_price - yesterday\_price}{yesterday\_price} * 100\% \qquad (1)$$

Python Data Frame function is used for this purposes $DataFrame.pct\_change(1)$

5.3 Normality Test

To execute normality test $stats.normaltest()$ function from Python scipy is used.

5.4 Lognormality Test

A random variable has a lognormal distribution if and only if logarithm of the variable follows normal distribution. So, the same function is used, but using logarithm values as argument.

5.5 Probability Test

The key point in the test is usage of $stats.norm.cdf(z_score)$.

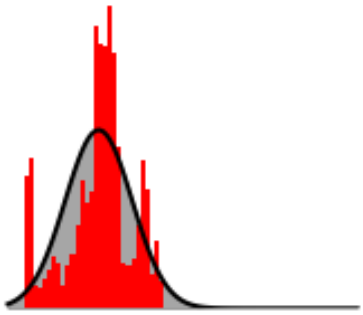
5.6 Visualization

The main visualization tool in my project is *matplotlib.pyplot* function $plt.subplots()$, which allows to plot several charts on the same figure. Two sets of graphs represent distribution of indices values (logarithms) and returns with corresponding theoretical normal curve.

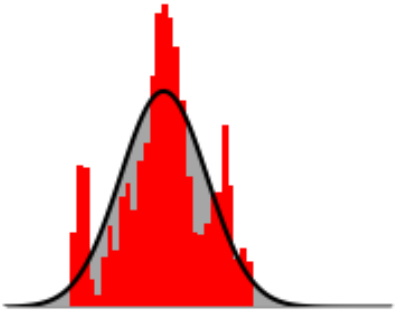
Distribution of indices logarithm values

^DJI, mu=9.20, sigma=0.45, kurt_pv=0.02

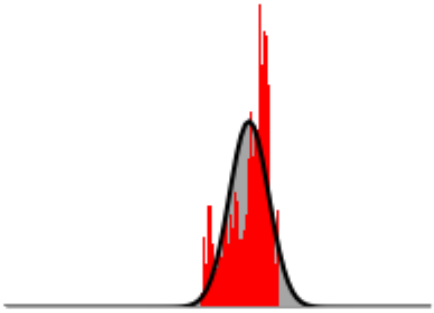
^GSPC, mu=7.04, sigma=0.44, kurt_pv=0.08



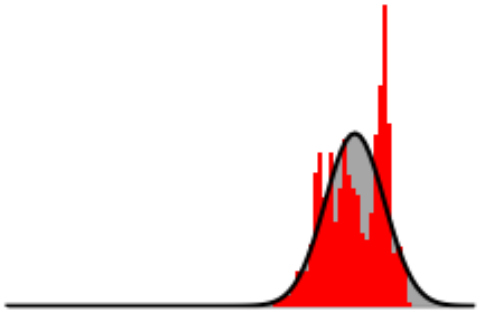
^IXIC, mu=7.68, sigma=0.56, kurt_pv=0.00



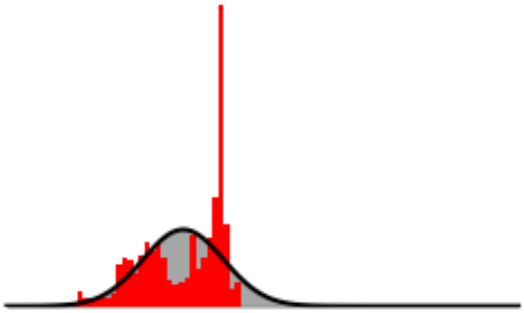
^FTSE, mu=8.54, sigma=0.25, kurt_pv=0.00



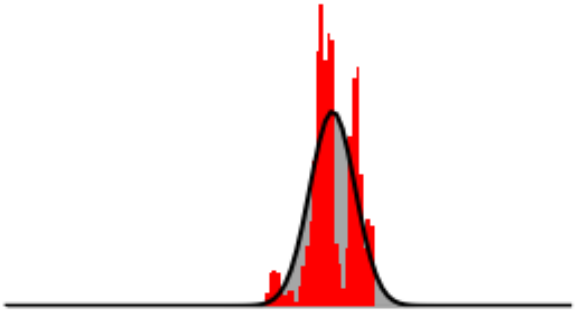
^HSI, mu=9.63, sigma=0.39, kurt_pv=0.00



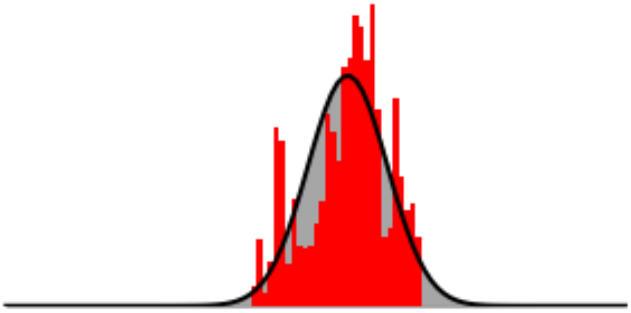
^KS11, mu=7.10, sigma=0.53, kurt_pv=0.00



^NSEI, mu=8.71, sigma=0.30, kurt_pv=0.20

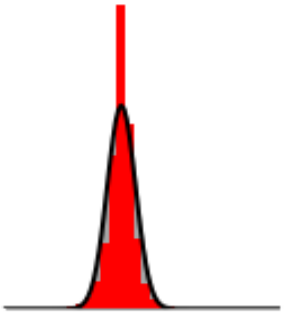


^GDAXI, mu=8.55, sigma=0.52, kurt_pv=0.00

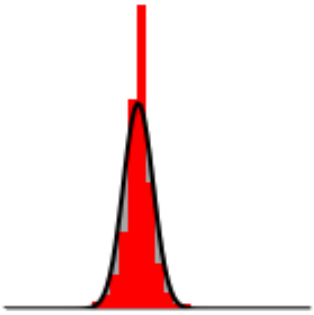


Distribution of indices returns

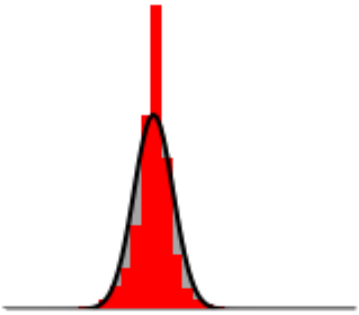
^DJI, mu=0.00, sigma=0.01, kurt_pv=0.00



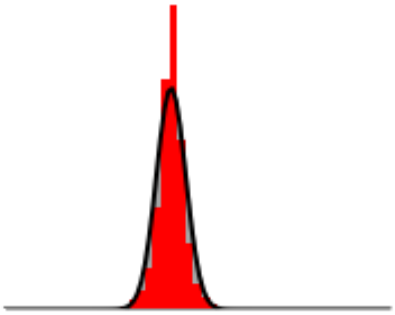
^GSPC, mu=0.00, sigma=0.01, kurt_pv=0.00



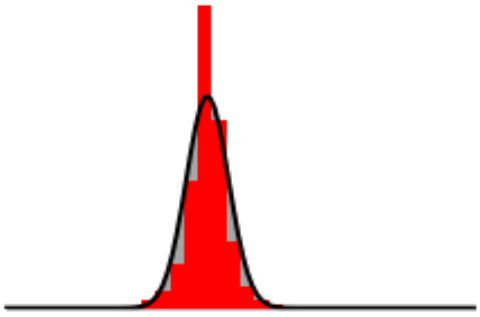
^IXIC, mu=0.00, sigma=0.01, kurt_pv=0.00



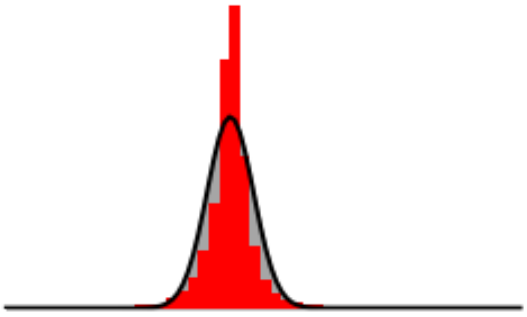
^FTSE, mu=0.00, sigma=0.01, kurt_pv=0.00



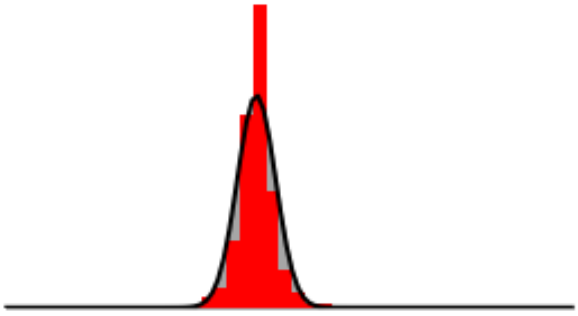
^HSI, mu=0.00, sigma=0.02, kurt_pv=0.00



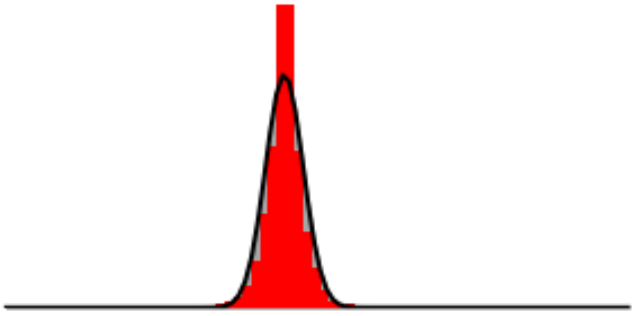
^KS11, mu=0.00, sigma=0.02, kurt_pv=0.00



^NSEI, mu=0.00, sigma=0.01, kurt_pv=0.00



^GDAXI, mu=0.00, sigma=0.01, kurt_pv=0.00



6 Conclusion

6.1 Price lognormality hypothesis

Based on the hypothesis test results for all indices one can conclude that all analyzed indices values do not follow lognormal distribution

^DJI	^GSPC	^IXIC	^FTSE
8.04777e-99	1.91724e-63	1.53789e-28	5.23048e-126
^HSI	^KS11	^NSEI	^GDAXI
1.22825e-256	5.6327e-141	1.24376e-14	5.20923e-75

So all the values are less than 0.05. This means that the lognormality distribution hypothesis is rejected.

6.2 Returns normality hypothesis

Based on the hypothesis test results for all indices one can conclude that all analyzed indices returns do not follow normal distribution.

^DJI	^GSPC	^IXIC	^FTSE
2.8946e-225	1.13041e-229	5.89787e-187	2.29292e-180
^HSI	^KS11	^NSEI	^GDAXI
1.15419e-263	1.47589e-130	2.37391e-129	8.56094e-145

So all the values are less than 0.05. This means that the normality distribution hypothesis is rejected.

6.3 Random walk test

The probability test showed that the probability of The Black Monday for Dow Jones Industrial Average is fairly zero, which means that such type of events are roughly impossible, however the practice shows the opposite.

6.4 Fat tails problem

The analyzed data has "Fat tails" that means that probability of theoretically unlikely events are considerably underestimated. It is clear from the

comparison represented on the pictures in section "Visualization"

Regarding "fat" tails, kurtosis test is used. Python function $stats.kurtosistest()$ gives the p-value. This function tests the null hypothesis that the kurtosis of the population from which the sample was drawn is that of the normal distribution.

Only 2 index values do not have "fat" tails: NSEI and GSPC